CS 221 Project Proposal

Team: Ambika Acharya(aacharya), Angela Sy(angelasy), Emily Tang(emjtang)

Research Question: Can we predict the star rating of a Yelp review given the review text?

Input: a single Yelp Review
Output: number of stars (1-5)

Scope: Yelp reviews for businesses in the Phoenix area
We have a substantial data set from the Yelp Academic Data Set with information on businesses and reviews.

Object Type	Count
Business	15,585
Reviews	335,022

Evaluation Metric: The percentage of review ratings we correctly predicted (prediction == ground truth label)

Example Set of Input and Output:
Input: I love Marilo! She understands my hair type and knows exactly what to do with my hair. She keeps a record of my previous visits. She recommends what is best for my hair. She is pleasant to work with: easygoing, friendly, and respectful. I've been going to her since 2008. I'm really picky with hair people, and I used to go back to Chicago for haircuts. Now, I stick to Marilo.
Output: 5 stars

Input: This is the first unique (not chain) brunch place I have found in the east valley that serves alcohol! The food was DELISH! finally
Output: 4 stars

Input: Lately i have been feeling homesick for asian food and been hitting up places that i haven't been to in awhile. Recently re-visited Thai Pan for a quick lunch and quickly ordered without spending too much time perusing the menu. It looked more diverse than I remembered including some Vietnamese additions. I remembered the curries and stir-fry dishes were ok but nothing really memorable. A quick summary for my latest visit:\n\nPros:\n- convenient order-at-the counter setup\n- self-serve drink station\n- brown and white rice mixture\n- friendly and gracious owners\n\nCons:\n- too much napa cabbage in comparison to green vegetables \n- wish the owner/chef would be back in the kitchen vs. managing\n- spice level on the weak side
Output: 3 stars

Input: Brian S. Sum's it up perfectly with \"if you're going to try to be a Hooters/Library sports bar rip-off, hire more attractive, more in-shape waitresses to wear those cheesy uniforms.\"\n\nThe only reason I give it a second star is because it's so close to home and they're open till 2, so I can't hate totally.
Output: 2 stars

Input: I went to this Petco a few times and was a bit appalled at their rodents section, to be perfectly honest. Granted, it wasn't as bad as it could have been, but in my opinion, it was unacceptable.\n\nFor one thing, there was six or seven rats packed in one cage that was, what . . . \n1' x 9\", 1 x 8\"? Don't quote me on that - I'm not

100% sure, but regardless of the details, it was way too small for the number of rats that were in it. They weren't unhealthy; they didn't look very sick, but they definitely didn't look particularly healthy, either. \n\nThe guinea pigs were what really shocked me, though. If you know anything about guinea pigs, you'll know they need a lot of space - even just for one. This cage was the same size as the rat's cage (they all were). They barely had enough to room to walk, much less exercise or popcorn (which my guinea pigs do often). There were two guinea pigs packed in that thing, when it was way, way too small for one baby guinea pig. It also didn't look that clean - all the cages, in my opinion, were grungy. The whole place felt grungy and grimy…Good prices? They don't make up for that.\n\nI apologize if I made this too long, and thank you if you read all of this, regardless of whether or not you agree with me.

Output: 1 star

Example Features
- word extractor (sentiment analysis)
- punctuation, capitalization
- number of characters (length of review → probably an extreme score like 1 or 5)

Baseline and Oracle

For our baseline algorithm, we modified the code from our sentiment analysis homework to extract word features from Yelp restaurant reviews and give them a score. Instead of producing a binary output (positive or negative), the baseline uses arbitrary thresholds to determine the number of stars (from 1 to 5) for a given review. The baseline uses the simplest feature extractor, taking the count of each word in each review, and giving a rating based on that.

Our oracle would be a human reading the review and giving a star rating. Humans are able to detect sarcasm, follow contradictions and deal with other challenges we list below, which a computer must be trained to do.

What are the challenges?
- Detecting sarcasm, for example: in one of the reviews above "Good prices?" The reviewer was being sarcastic and ended up giving the restaurant 1 star only.
- Differentiating between 4 and 5 stars, or 1 and 2 stars (What warrants extreme ratings?)
- Dealing with varying lengths of reviews (one review could be a sentence, another 5 paragraphs)

Topics that address the challenges
- Good feature extraction and classification
- Supervised machine learning
- Sentiment analysis
- Natural Language Processing

Related Work
- Prediction of Helpful Reviews using Emotions Extraction
  http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/viewFile/8562/8607
- Predicting Yelp Restaurant Reviews
  http://www.seas.upenn.edu/~chuyaguo/pdf/HungryHubbleAnalysis.pdf
- Beyond the Stars: Improving Rating Predictions using Review Text Content
  http://spidr-ursa.rutgers.edu/resources/WebDB.pdf
- Identifying Sarcasm in Twitter: A Closer Look
  http://hnk.ffzg.hr/bibl/acl2011/Short/pdf/ACL-HLT2011102.pdf